

シソーラス情報を用いたソフトウェア開発の支援

著者	早瀬 康裕
発行年	2012
その他のタイトル	Supporting Software Development Using Thesaurus
URL	http://hdl.handle.net/2241/118691

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 3月31日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2009～2011

課題番号：21700031

研究課題名（和文）シソーラス情報を用いたソフトウェア開発の支援

研究課題名（英文）Supporting Software Development Using Thesaurus

研究代表者

早瀬 康裕（HAYASE YASUHIRO）

筑波大学・システム情報系・助教

研究者番号：40423090

研究成果の概要（和文）：

プログラム理解では識別子の名前が非常に重要な手がかりとなることが知られている。本研究では、ソースコードから命名の事例を収集することで、名詞の上位・下位関係と、動詞と名詞の動詞・目的語関係、名詞の説明文が収録された辞書を自動的に生成する手法を提案した。さらに、作成した辞書を利用し、ソフトウェア開発者が新しく作ろうとしているメソッド名の候補を生成し、より適切な命名をうながすシステムを開発した。

研究成果の概要（英文）：

In program comprehension, the names of identifiers in target program is one of the biggest key to comprehension. This project proposed a system to build a dictionary by collecting cases of identifier names. And then, the dictionary is provided for software developers to support program identifier naming.

交付決定額

（金額単位：円）

	直接経費	間接経費	合 計
2009 年度	1,300,000	390,000	1,690,000
2011 年度	1,200,000	360,000	1,560,000
2012 年度	700,000	210,000	910,000
総 計	3,200,000	960,000	4,160,000

研究分野：ソフトウェア工学

科研費の分科・細目：情報学、ソフトウェア

キーワード：辞書、識別子、プログラム理解、命名支援

1. 研究開始当初の背景

ソフトウェア保守にかかるコストはソフトウェアの総所有コストのうち大半を占めることが知られており、その中でもソフトウェア理解にかかるコストが半分以上であるという報告がある。保守の過程でソフトウェアを変更するためには、対象ソフトウェアを詳細に理解しなければならない。しかし、ソフトウェアに付随するべきドキュメントは、ソフトウェア保守の過程で適切に変更されないことが多い[1] ため、ソフトウェアを理

解するにはソースコードを読み、理解しなければならない[2]。

ソースコードを理解する際に、保守作業者は様々な手がかりを利用するが、ソースコード中に出現する識別子の名前は非常に重要な手がかりの1つであるとされている。保守作業者がソースコードを理解する時には、そのプログラムの使われるドメインの知識とプログラム要素との対応付けに識別子の名前を用いる[3][4]。このため、識別子に適切な名前が付けられていない場合には、ソースコードを理解するために莫大な時間を必要

としてしまう[5]。さらに、識別子に適切な名前が付けられていたとしても、作業者が保守対象のソフトウェアに不慣れな場合にはプログラム要素と知識との対応付けが上手く出来ず、熟練者に比べて保守作業に長い時間がかかってしまう。

一方、効率的にソフトウェアを生産する手法の1つにソフトウェアの再利用がある。開発者は、再利用可能なソフトウェア部品を得るために、キーワード検索を用いる。しかし、適切なキーワードを思いつかない場合には必要な部品を得ることが出来ず、生産性を改善できない。

申請者の研究グループではソフトウェア部品検索システムを開発しており、この一環として、カテゴリ検索を検索システムに組込んだ[6]。このカテゴリ階層を構築するために、検索システムに登録された Java ソースコードから、単語間の上位下位関係(抽象-具象関係)と修飾関係を表したシソーラス(単語間の関連をグラフとして記録した辞書)を作成した。

作成したシソーラスには自然言語のシソーラスには存在しない関係が含まれており、これを保守作業者に提供することでソフトウェア理解に役立てることが出来るのではないかと考えた。しかし、申請者らが開発したシソーラス構築システムには不適切な上位下位関係が全体の約 1/3 存在するという問題があった。この原因は、不適切な名前が付けられた識別子がソースコード中に存在し、それをシソーラスに登録してしまったことである。

2. 研究の目的

本研究では、ソフトウェア開発の様々な場面で開発者を支援するために、ソフトウェア中で用いられる単語の辞書を自動的に作成し、開発者が必要とする情報を提供することを目的とする。作成する辞書は単語間の関係をグラフ構造で表したデータ(シソーラス)であり、識別子に用いられる単語と、その単語間の関係が記録されている。

具体的には、前述のシソーラスの精度を改善する一方で、上位下位関係や修飾関係以外の有用な情報を抽出する。さらに、作成したシソーラスを既存の自然言語のシソーラス(WordNet 等)と比較することで、ソフトウェアに固有の単語を抽出することを目指す。また、作成したシソーラスを用いることで、ソフトウェアを効率的に開発および保守するシステムを作成する。ソフトウェア設計者と保守作業者には、作業内容に応じた情報を提供することで、識別子の命名やソフトウェア理解を支援する。ソフトウェア検索システム

には、検索条件として与えるキーワード(検索クエリ)をシソーラス上の関連する単語を用いて拡張することで、検索結果の再現率を高めることを目指す。

このようなシソーラスを作成した研究は研究開始時には存在しなかった。また、このシソーラスを用いることで、識別子に適切な命名を行なうことや、開発されるソフトウェアの品質向上に寄与するものと考えられる。また、辞書を活用することで、ソフトウェア検索システム上でより柔軟な検索が行えるようになり、ソフトウェア再利用を活性化することができると考えられる。

3. 研究の方法

3.1 シソーラスの改良

シソーラスに記録する関係や単語の種類を増やす。既に取得している上位下位関係と修飾関係に加えて、動詞と目的語の関係や、単語の説明を記録する。動詞と目的語の関係については、メソッド名と呼び出される側のオブジェクトの名前や型名から間接目的語の関係を抽出し、メソッド名と引数から直接目的語の関係を抽出する。単語の説明を生成する方法は、クラスやメソッド、変数などを説明するコメント(ドキュメントコメント)を自然言語の解析手法を用いて要約することで行う。

2つ目の改良として、シソーラスから不適切な単語や関係を除去する。ソースコード中に現れる識別子は一律に注意して名前が付けられておらず、一般に、モジュール外部から利用される識別子は熟慮して名付けられるのに対し、モジュール内部のみで使用される識別子の名前にはあまり注意が払われない。これを利用し、ソースコード内からのみ参照可能な識別子や、外部からの参照が少ない識別子から取得した単語間の関係は、同じ関係が複数のソースコードで出現した時のみシソーラスに登録するようシステムを変更する。外部からの参照の多さの基準としては、fan-in やコンポーネントランク値(実績に基く再利用性の尺度)[9]を用いる。

3.2 開発者へのシソーラス情報の提供

統合開発環境を拡張することで、シソーラスに記録された情報を、開発者に提供するシステムを開発する。具体的には、識別子への命名に悩んでいる開発者に対して、シソーラスを活用して識別子名の候補を提示する機能を、統合開発環境に追加する。

開発した機能の有効性を検証するため、被験者実験を行い、機能を利用することで識別

子への命名がどのように変化するかを調べる。

3.3 ソフトウェア部品検索システムでのシソーラスの利用

キーワード検索に基づくソフトウェア部品検索システム SPARS を拡張し、シソーラス等を活用することで、より柔軟で効率的な検索が行えるようにする。

4. 研究成果

研究計画に対応する成果について、その内容と対応する発表を述べる。

まず、シソーラスに格納する情報を増やす研究の一環として、プログラム中に出現する動詞-目的語関係を取得し、辞書に収録するシステムを開発した。この研究については、学会発表(11)で計画を示し、学会発表(4)(8)において成果を報告した。

また、同じくシソーラスに格納する情報を増やす研究の一環として、プログラム中に出現する名詞の説明を生成し、辞書に収録するシステムを開発した。この研究については、学会発表(10)で計画を示し、学会発表(7)で成果を報告した。

次に、シソーラス情報をソフトウェア開発者に提供する研究として、オブジェクト指向プログラミング言語で書かれたソースコードに対して、新規にメソッドを作成しようとしている開発者に対して、メソッド名の候補を提示するシステムを開発した。このシステムについては、学会発表(1)でその成果を公表した。

さらに、シソーラス情報を活用した、ソフトウェア検索システムの効率的な利用については、その構想を拡張し、学会発表(3)において発表することで関連分野の研究者から意見を収集した。

参考文献

- [1] T. D. LaToza, G. Venolia, R. DeLine, "Maintaining mental models: a study of developer work habits", 28th international conference on software engineering, pp.492-501, 2006.
- [2] T. D. LaToza, D. Garlan, J. D. Herbsleb, B. A. Myers, "Program comprehension as fact finding", 6th ESEC/FSE, pp361-370, 2007.
- [3] A. V. Mayrhauser, A. M. Vans. "Identification of dynamic comprehension processes during largescale maintenance". IEEE transactions on software engineering,

Vol. 22, No. 6, pp424-437, 1996.

[4] N. Pennington. "Comprehension strategies in programming", Empirical studies of programmers: second workshop, pp.100-113, 1987.

[5] D. Lawrie, C. Morrel, H. Feild, D. Binkley, "What's in a name? a study of identifiers". In proc. of the 14th international conference on program comprehension, pp.3-12, 2006.

[6] 早瀬 康裕, 市井 誠, 井上 克郎, "ソフトウェア理解支援を目的とした辞書の作成法", ウィンターワークショップ 2008・イン・道後論文集, pp.33-34, 2008.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 11 件)

- (1) 鬼塚 勇弥, 早瀬 康裕, 石尾 隆, 井上 克郎: "ソースコード中に出現する動詞-目的語関係を利用したメソッド名の命名支援手法", 電子情報通信学会ソフトウェアサイエンス研究会, 2012/3/13, てんぶす那覇, 沖縄県
- (2) 中村 高士, 早瀬 康裕, 北川 博之: "プロジェクト横断的なオープンソースソフトウェア開発記録の分析手法", 情報処理学会第 74 回全国大会, 2012/3/7, 名古屋工業大学, 愛知県
- (3) 駒水 孝裕, 早瀬 康裕, 北川 博之: "ソフトウェア部品検索に適したファセット探索の一考察", 電子情報通信学会ソフトウェアサイエンス研究会, 2011 年 10 月 27 日、北陸先端科学技術大学院大学, 石川県
- (4) Yasuhiro Hayase, Yu Kashima, Yuki Manabe, Katsuro Inoue: "Building Domain Specific Dictionaries of Verb-Object Relation from Source Code", 15th European Conference on Software Maintenance and Reengineering (CSMR 2011), Oldenburg, Germany, March 1-4, 2011
- (5) Yu Kashima, Yasuhiro Hayase, Norihiro Yoshida, Yuki Manabe and Katsuro Inoue: "An Investigation into the Impact of Software Licenses on Copy-and-Paste Reuse among OSS Projects", 18th Working Conference on

Reverse Engineering (WCRE 2011),
Limerick, Ireland, October 17-20, 2011

(2)研究分担者
なし

- (6) Yu Kashima, Yasuhiro Hayase, Norihiro Yoshida, Yuki Manabe, Katsuro Inoue: "A Preliminary Study on Impact of Software Licenses on Copy-and-Paste Reuse", International Workshop on Empirical Software Engineering in Practice 2010 (IWESEP 2010), Nara, Japan, December 7-8, 2010

(3)連携研究者
なし

- (7) 藤木 哲也, 早瀬 康裕, 井上 克郎: "コメント文によるプログラム中に出現する名詞の説明文生成", 電子情報通信学会ソフトウェアサイエンス研究会, 2010/8/6, 旭川市民文化会館, 北海道

- (8) 鹿島 悠, 早瀬 康裕, 真鍋 雄貴, 松下 誠, 井上 克郎: "メソッドに用いられる動詞-目的語関係を収録した辞書構築手法の提案", 情報処理学会ソフトウェア工学研究会, 2010/6/2, 化学会館, 東京都

- (9) Makoto Higashi, Tetsuo Yamamoto, Yasuhiro Hayase, Takashi Ishio, Katsuro Inoue: "An Effective Method to Control Interrupt Handler for Data Race Detection", 5th International Workshop on Automation of Software Test (AST 2010), Cape Town, South Africa, May 4, 2010

- (10) 藤木 哲也, 早瀬 康裕, 井上 克郎: "識別子の読解を目的とした名詞辞書の作成方法の一試案", ウィンターワークショップ 2010・イン・倉敷, 情報処理学会シンポジウムシリーズ, 2010/1/21, 倉敷市芸文館, 岡山県

- (11) 鹿島 悠, 早瀬 康裕, 井上 克郎: "メソッド名に用いられる動詞-目的語関係の辞書作成にむけて", ウィンターワークショップ 2010・イン・倉敷, 情報処理学会シンポジウムシリーズ, 2010/1/21, 倉敷市芸文館, 岡山県

6. 研究組織

(1)研究代表者

早瀬 康裕 (HAYASE YASUHIRO)
筑波大学・システム情報系・助教
研究者番号: 40423090